TRANSMITTAL FORM

(to be used for all correspondence after initial filing)

Application Number	09/835,064
Filing Date	July 11, 2005
First Named Inventor	Giuffrida
Art Unit	2175
Examiner Name	Neveen, Abel-Jalil
Attorney Docket Number	HRL065

Total Number of Pages in This Submission **52**

ENCLOSURES *(Check all that apply)*

- [✓] Fee Transmittal Form
  - [✓] Fee Attached
- [ ] Amendment/Reply
  - [ ] After Final
  - [ ] Affidavits/declaration(s)
- [ ] Extension of Time Request
- [ ] Express Abandonment Request
- [ ] Information Disclosure Statement
- [ ] Certified Copy of Priority Document(s)
- [ ] Reply to Missing Parts/ Incomplete Application
  - [ ] Reply to Missing Parts under 37 CFR 1.52 or 1.53

- [ ] Drawing(s)
- [ ] Licensing-related Papers
- [ ] Petition
- [ ] Petition to Convert to a Provisional Application
- [ ] Power of Attorney, Revocation Change of Correspondence Address
- [ ] Terminal Disclaimer
- [ ] Request for Refund
- [ ] CD, Number of CD(s) _____
  - [ ] Landscape Table on CD

Remarks

- [ ] After Allowance Communication to TC
- [ ] Appeal Communication to Board of Appeals and Interferences
- [✓] Appeal Communication to TC **(Appeal Notice, Brief, Reply Brief)**
- [ ] Proprietary Information
- [ ] Status Letter
- [✓] Other Enclosure(s) (please Identify below):
  Return postcard.

SIGNATURE OF APPLICANT, ATTORNEY, OR AGENT

Firm Name	Tope-McKay & Associates
Signature	
Printed name	Cary Tope-McKay
Date	July 11, 2005
Reg. No.	41,350

CERTIFICATE OF TRANSMISSION/MAILING

I hereby certify that this correspondence is being facsimile transmitted to the USPTO or deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on the date shown below:
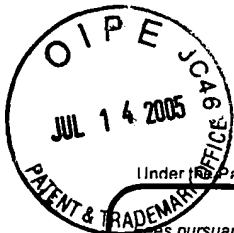
Signature	
Typed or printed name	Cary Tope-McKay
Date	July 11, 2005

Effective on 12/08/2004.
Fees pursuant to the Consolidated Appropriations Act, 2005 (H.R. 4818).

FEE TRANSMITTAL
For FY 2005

☐ Applicant claims small entity status. See 37 CFR 1.27

TOTAL AMOUNT OF PAYMENT	($)	500.00

Complete if Known

Application Number	09/835,064
Filing Date	04/13/2001
First Named Inventor	Giuffrida
Examiner Name	Neveen, Abel-Jalil
Art Unit	2175
Attorney Docket No.	HRL065

METHOD OF PAYMENT (check all that apply)

☐ Check ☑ Credit Card ☐ Money Order ☐ None ☐ Other (please identify): _____

☐ Deposit Account   Deposit Account Number: 50-2691   Deposit Account Name: _____

For the above-identified deposit account, the Director is hereby authorized to: (check all that apply)

☑ Charge fee(s) indicated below

☐ Charge fee(s) indicated below, **except for the filing fee**

☑ Charge any additional fee(s) or underpayments of fee(s) under 37 CFR 1.16 and 1.17

☑ Credit any overpayments

WARNING: Information on this form may become public. Credit card information should not be included on this form. Provide credit card information and authorization on PTO-2038.

FEE CALCULATION

1. BASIC FILING, SEARCH, AND EXAMINATION FEES

Application Type	FILING FEES Fee ($)	Small Entity Fee ($)	SEARCH FEES Fee ($)	Small Entity Fee ($)	EXAMINATION FEES Fee ($)	Small Entity Fee ($)	Fees Paid ($)
Utility	300	150	500	250	200	100	_____
Design	200	100	100	50	130	65	_____
Plant	200	100	300	150	160	80	_____
Reissue	300	150	500	250	600	300	_____
Provisional	200	100	0	0	0	0	_____

2. EXCESS CLAIM FEES

Fee Description	Fee ($)	Small Entity Fee ($)
Each claim over 20 (including Reissues)	50	25
Each independent claim over 3 (including Reissues)	200	100
Multiple dependent claims	360	180

Total Claims	Extra Claims	Fee ($)	Fee Paid ($)
_____ - 20 or HP =	_____ x	_____ =	_____

HP = highest number of total claims paid for, if greater than 20.

Indep. Claims	Extra Claims	Fee ($)	Fee Paid ($)
_____ - 3 or HP =	_____ x	_____ =	_____

HP = highest number of independent claims paid for, if greater than 3.

Multiple Dependent Claims

Fee ($)	Fee Paid ($)
_____	_____

3. APPLICATION SIZE FEE

If the specification and drawings exceed 100 sheets of paper (excluding electronically filed sequence or computer listings under 37 CFR 1.52(e)), the application size fee due is $250 ($125 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

Total Sheets	Extra Sheets	Number of each additional 50 or fraction thereof	Fee ($)	Fee Paid ($)
_____ - 100 =	_____ / 50 =	_____ (round up to a whole number) x	_____ =	_____

4. OTHER FEE(S)

Non-English Specification, $130 fee (no small entity discount)

Other (e.g., late filing surcharge): Fee to file an Appeal Brief _____

Fees Paid ($)

$ 500.00

SUBMITTED BY

Signature	Registration No. (Attorney/Agent) 41,350	Telephone (310) 589-8158
Name (Print/Type) Cary Tope-McKay		Date July 11, 2005

O I P E JUL 1 4 2005 PATENT & TRADEMARK OFFICE

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

5

App. No.   :   09/835,064

Applicants   :   Giuffrida et al.

Filed   :   04/13/2001

10   TC/A.U.   :   2175

Examiner   :   Neveen, Abel-Jalil

Docket No. :   HRL065

15

Customer No. :   28848

For   :   "A Method and Apparatus for Automatically Extracting Metadata from
                    Electronic Documents Using Spatial Rules."

20

BRIEF ON APPEAL

25   Hon. Commissioner for Patents
     Washington, D.C. 20231

Sir:

       This is an appeal from the Final Rejection, dated January 12, 2005, for the above

30   identified patent application.

REAL PARTY IN INTEREST

       The present application has been assigned to HRL Laboratories, LLC of Malibu,

CA.

35   ## RELATED APPEALS AND INFERENCES

       There are no related appeals or interferences to this application.

                          ## STATUS OF CLAIMS

       Claims 1-16 are the subject of this appeal. A copy of all claims of the application

is contained in the attached Appendix A.

STATUS OF AMENDMENTS

No Amendment after Final Rejection has been entered.

SUMMARY OF INVENTION

5   The invention described and claimed in the present application relates to a method and apparatus for extracting metadata from electronic documents. More specifically, the invention described and claimed in the present application relates to a method and apparatus for combining text-based matching and spatial reasoning used in the extraction of metadata. In one embodiment, spatial knowledge-based methodology is used for

10   document disaggregation. In one embodiment, visual and spatial knowledge are exploited when reading a document. In general, within a document category, a certain visual layout can be identified for all documents within that category. For instance, a scientific paper may follow the format described below, wherein the uppercase words represent metadata in the paper and bold words denote spatial relationships and other

15   types of relationships.

The TITLE is located on the **upper** portion of the **first page** and it is printed using the

largest font on the **first page**;

AUTHORS are listed **immediately under** the TITLE in some order;

20   AFFILIATIONS **follow** the authors' list;

If **only one** AFFILIATION appears then **all** AUTHORS are associated with it;

The same **font** is used for all AUTHORS and, similarly, for all AFFILIATIONS;

The FIRST LEVEL HEADERS use a larger font than the SECOND LEVEL headers.

In one embodiment of the present invention, rule-based language is used to encode the

25   visual layout of the document. Different types of documents require different knowledge

bases. A knowledge base is encoded with visual and spatial layout facts. The knowledge

base described in the example above deals with scientific papers appearing in conference proceedings and specialized journals.

The metadata extraction system retains the document's original formatting. Formatting

5   includes both font size and text positioning on the page. Hereinafter, data that retains the original document's formatting shall be referred to as substantially format-invariant data.

Electronic documents go to an intermediate language conversion step, which converts the electronic documents into substantially format-invariant data files, and captures the

10   spatial and visual aspects for document representation. This can generally be achieved by transferring the original document to a file from the default viewer of the document. A converted document has to undergo a spatial layout fact extraction process to extract relevant spatial layout information and eliminate irrelevant information from the converted document in preparation for further processing. This is a task generally

15   accomplished by any substantially format-invariant data printer driver or viewer.

After spatial layout facts have been extracted from a substantially format-invariant data file, spatial layout facts are subjected to spatial metadata reasoning. A knowledge engineer provides a set of spatial layout rules that embodies the protocol for extracting the metadata of interest from the provided document. A rule-based language is used to

20   read the provided format-invariant data file and to produce a set of spatial layout facts for the rule-based language. Each fact contains information—text and spatial data—about the inputed substantially format-invariant data document. Rules provided by the

knowledge engineer reason with the extracted facts to identify and extract relevant

metadata from the input documents.

ISSUES

Issue 1 – Are Claims 1-5, 7-13 and 15-16 patentable under 35 USC 102(b) over

5   U.S. Patent No. 5,920,856 to Syeda-Mahmood ("the Syeda-Mahmood patent")?

GROUPING OF CLAIMS

The ground of rejection which appellant contests herein applies to more than one

claim, such additional claims, to the extent separately identified and argued below, do not

stand or fall together.

10   **THE ARGUMENT**

Issue 1 – Are Claims 1-5, 7-13 and 15-16 patentable under 35 USC 102(b) over U.S.
Patent No. 5,920,856 to Syeda-Mahmood ("the Syeda-Mahmood patent")?

In section 3 of the Office Action of January 12, 2005, the Examiner rejected

Claims 1-5, 7-13 and 15-16 under 35 USC §102(b) as being unpatentable over U.S.

15   patent No. 5,920,856 to Syeda-Mahmood, herein referred to as the "Syeda-Mahmood

patent." The Applicants submit that the Syeda-Mahmood patent does not describe each

and every element as set forth in the rejected claims.

20   Claim 1

First, the Applicants wish to reiterate the position stated in the Applicants'

response after final rejection dated January 12, 2005, that the Examiner has incorrectly

defined the preamble of Claim 1. The preamble of Claim 1 states, "[a]n apparatus for

automatically extracting metadata from electronic documents." In the first sentence of

25   the second full paragraph on page 6 of the referenced Office Action, the Examiner stated

that he "refers to figure 1 to show the preamble of the claim." The Applicants assert that

the Examiner was mistaken as to what is contained in the preamble. The Applicants
submit that the preamble ends with the term before the transition word "comprising."
Thus, the preamble of Claim 1 states, "[a]n apparatus for automatically extracting
metadata from electronic documents." Further, the Applicants submit that FIG. 1 of the
5   Syeda-Mahmood patent does not disclose an apparatus for automatically extracting
metadata from electronic documents.

The Applicants submit that for many of the claims rejected under 35 USC 102(b),
the Examiner reached an improper conclusion as to the teachings of the Syeda-Mahmood
patent and incorrectly interpreted the limitations of the rejected claims. In order to
10   establish a *prima facie* case of anticipation, the Examiner must set forth an argument that
provides (1) a single reference (2) that teaches or enables (3) each of the claimed
elements (as arranged in the claim) (4) either expressly or inherently, and (5) as
interpreted by one of ordinary skill in the art. All of these factors must be present, or a
case of anticipation is not met. Thus, "[a]nticipation requires that every element of the
15   claims appear in a single reference ..." Continental Can Co. v. Monsanto Co. 948 F.2d
1264 (Fed. Cir. 1991). The Applicants assert that the Syeda-Mahmood patent does not
describe each and every element as set forth in the rejected claims.

Regarding Claim 1, the Examiner asserted, in section 3 of the Final Office Action,
that figure 1 of the Syeda-Mahmood patent discloses an apparatus, and a method for
20   automatically extracting metadata from electronic documents comprising a first
processing element, a second processing element, a reasoning element, and a database.
The Examiner stated that col. 7, lines 1-39, and col. 12, lines 47-55 of the Syeda-
Mahmood patent discloses a first processing element configured to convert electronic
documents into files. The Examiner further stated that col. 7, lines 1-39 of the Syeda-
25   Mahmood patent discloses that said first processing element is configured to provide the
files to a second processing element. The Examiner interpreted the translator module of
the Syeda-Mahmood patent that resides in the meta-database to be equivalent to the first
processing element, since the translator module converts electronic documents from
different Web sites into records stored at a central location. The Examiner also stated
30   that col. 7, lines 1-45 of the Syeda-Mahmood patent teaches that said second processing
element is configured to recieve files and extract predetermined information. In addition,

the Examiner stated that col. 7, lines 1-45 of the Syeda-Mahmood patent teaches that said second processing element is further configured to provide said extracted predetermined information to said reasoning element. The Examiner asserted that the search agent of the Syeda-Mahmood patent reads on the second processing element of the present

5 application, and metadata of the Syeda-Mahmood patent reads on predetermined information of the present application. Further, the Examiner stated that col. 7, lines 15- 38 of the Syeda-Mahmood patent discloses that said database is configured to also provide input to said reasoning element. The Examiner also stated that col. 6, lines 1-22 and col. 7, lines 14-38 of the Syeda-Mahmood patent discloses that said reasoning

10 element is configured to use a set of rules to extract metadata from the files. Finally, the Examiner stated that col. 5, lines 24-38 of the Syeda-Mahmood patent discloses that said reasoning element provides an output of metadata, where the Examiner asserted that the reasoning element of the present application reads on the refining module of the Syeda- Mahmood patent.

15   The Applicants disagree with the characterization of the Syeda-Mahmood patent by the Examiner. First, the Applicants point out that the system disclosed in the Syeda- Mahmood patent is very different than the system claimed in the present application. Looking to the abstract, the Syeda-Mahmood patent discloses a network server which interfaces a client with selected database sites from a plurality of database sites. The

20 network server comprises a meta-database, a search agent, and a refining module. The search agent indexes the meta-database with a user query obtained from the client, and then distributes queries, developed pursuant to such indexing, to the selected ones of the plurality of database sites. A refining module is used to update the meta-database with the database relevancy information. The Applicants submit that this is very different

25 from an apparatus for automatically extracting metadata from electronic documents, wherein said second processing element is configured to provide extracted predetermined information to a reasoning element, and the reasoning element is configured to use a set of rules to extract metadata from the files.

        The Applicants look to the elements as arranged in Claim 1. Claim 1 claims, in

30 relevant part, a first processing element configured to convert electronic documents into files, a second processing element configured to receive the files and extract

predetermined information, and a reasoning element configured to use a set of rules to

extract metadata from the files, where the reasoning element provides an output of

metadata. In the Examiner's rejection of Claim 1 on pages 2 and 3 of the referenced

Office Action, the Applicants understand that the Examiner has analogized the first

5   processing element to the translator module of the Syeda-Mahmood patent, the second

processing element to the search agent of the Syeda-Mahmood patent, the predefined

information to the metadata of the Syeda-Mahmood patent, and the reasoning element to

the refining module of the Syeda-Mahmood patent. However, in the same referenced

Office Action on page 6, the Examiner appears to have changed his mind, asserting that

10   the second processor is not anticipated by the search agent of the Syeda-Mahmood patent,

but instead by the translator module of the Syeda-Mahmood patent. However,

previously, the Examiner had asserted that the first processing element was the translator

module. Additionally, on page 6, the Examiner stated that a search agent of the Syeda-

Mahmood patent is capable of being a reasoning element; however, this contradicted the

15   Examiner's earlier assertion that the refining module of the Syeda-Mahmood patent

anticpates the reasoning element. Given the Examiner's conflicting positions, the

Applicants are confused as to which elements of the Syeda-Mahmood patent the

Examiner asserts anticipate which elements of the present claims.

         If the Examiner is correct and the various components of the Syeda-Mahmood

20   patent anticipate the elements in Claim 1, then, following the Examiner's logic, the

Syeda-Mahmood patent must teach that the translator module converts electronic

documents into files, the search agent receives the converted files and uses query patterns

to extract predetermined information or the search agent receives the converted files and

extracts metadata, and the refining module or the search agent uses a set of rules to

25   extract metadata from the files, where the refining module provides an output of

metadata. The Applicants submit that the Syeda-Mahmood patent does not teach

embodiments that operate in the manner described above. Instead, the Applicants submit

that the Syeda-Mahmood patent teaches a method of querying databases and web-sites

for information, storing and updating that information, and then retrieving that

30   information at a later time.

First, the Applicants note that the Examiner cannot claim that the search agent of the Syeda-Mahmood patent is both the second processing element and the reasoning element. If the Examiner states that the search agent is the second processing element, and the search agent extracts metadata (i.e., predetermined information), then how does the

5   Examiner anticipate the reasoning element is configured to use a set of rules to extract metadata from files? The Applicants submit that the Syeda-Mahmood patent does not disclose all of the elements of Claim 1.

Col. 7, lines 15-17 of the Syeda-Mahmood patent states that the user query may have to be transformed by a translator module into forms the respective databases expect.

10   Thus, the Syeda-Mahmood patent teaches that the translator module changes a query from one query language to another query language, much like a Spanish translator will translate Spanish into English. The Applicants submit that this is not the same thing as "converting electronic documents into files." Thus, the Applicants submit that the Syeda-Mahmood patent does not teach, disclose or suggest a first processing element configured

15   to convert electronic documents into files, as is claimed in Claim 1.

Col. 6, lines 30-35 of the Syeda-Mahmood patent state that the role of the search agent 5 is to not only distribute queries to relevant database sites 8 after indexing into the meta-database 4, but also to assemble and return the database responses back to the user 6. The Applicants submit that this is not the same thing as "said second processing

20   element is configured to receive files and extract predetermined information." The Applicants submit that the Syeda-Mahmood patent teaches that the search agent distributes queries. Thus, the search agent requests predetermined information, but it does not receive files and extract predetermined information as is claimed in Claim 1.

Col. 5, lines 33-36 of the Syeda-Mahmood patent state that the refining module 7

25   (also referred to as training or learning module) associates patterns of responses from the databases to queries to learn the relevance of database sites 8 for similar further queries. Further, col. 6, lines 33-36 state that the search agent 5 relays the query data patterns and the database responses to the refining module 7 for site and scope relevancy updation. The Applicants do not know where in the Syeda-Mahmood patent the Examiner has

30   found support for the refining module 7 extracting metadata from the files, and outputting the metadata. For example, col. 7, lines 36-39 state that the output of the refining module

is a probability value indicating the likelihood of a database returning useful results for a given query. Thus, the Applicants submit that the Syeda-Mahmood patent does not teach, disclose or suggest "said reasoning element is configured to use a set of rules to extract metadata from the files; and said reasoning element provides an output of

5   metadata," as is claimed in Claim 1.

Since the Examiner has been unable to adequately show that the Syeda-Mahmood patent discloses several of the elements of Claim 1, the Applicants submit that the Examiner has not established a *prima facie* case for the rejection of Claim 1 under 35 USC 102. Further, as discussed above, the Applicants submit that the Syeda-Mahmood

10   patent does not teach, disclose, or suggest these limitations. Since the Syeda-Mahmood patent does not teach each and every element as set forth in Claim 1, Claim 1 is patentable for the reasons set forth above.

Claim 2

15   Claim 2, dependent on Claim 1, is patentable by virtue of its dependency.

Claim 3

Claim 3, dependent on Claim 1, is patentable by virtue of its dependency.

The Applicants further submit that Claim 3 is also patentable on its own merits.

20   In his rejection of Claim 3, the Examiner stated that the Syeda-Mahmood patent discloses an apparatus for automatically extracting metadata from electronic documents, wherein said predetermined information is substantially spatial layout facts, pointing to col. 9, lines 51-67 and col. 10, lines 15-36. The Applicants respectfully disagree with the conclusion drawn by the Examiner.

25   First, as noted above, the Examiner has indicated that the predetermined information is analogous to the query template or metadata in the Syeda-Mahmood patent. Col. 9, lines 51-67 deal with image feature extraction, which includes method for extracting information about edges, curves, color, texture, etc. Col. 10, lines 15-36 deal with image organization, specifically, clustering of image data and spatial information.

30   However, it is important to refer back to Claim 1 to understand Claim 3. Claim 1 claims, in part, "said second processing element is configured to receive said files and extract

predetermined information, said second processing element is further configured to

provide said extracted predetermined information to said reasoning element." Claim 3

claims that said predetermined information is substantially spatial layout facts. Thus, in

order for the Syeda-Mahmood patent to anticipate Claim 3, it must teach that said second

5   processing element is configured to receive said files and extract substantially spatial

layout facts, and that said processing element is further configured to provide said

extracted substantially spatial layout facts to said reasoning element. Col. 9, lines 51-67

discloses extracting information about images such as edges, curves, color and texture.

The Applicants submit that this does not anticipate that said second processing element is

10   configured to receive said files and extract substantially spatial layout facts, and that said

processing element is further configured to provide said extracted substantially spatial

layout facts to said reasoning element of Claim 3. Further, col. 10, lines 16-36 teaches

clustering image and spatial information in order to prevent false matches during

indexing. The Applicants submit that col. 10, lines 16-36 do not teach, disclose or

15   suggest that said second processing element is configured to receive said files and extract

substantially spatial layout facts, and that said processing element is further configured to

provide said extracted substantially spatial layout facts to said reasoning element of

Claim 3. Thus the Applicants submit that Claim 3 is patentable over the cited prior art.

20   Claim 4

         Claim 4, dependent on Claim 1, is patentable by virtue of its dependency.

     Claim 5

         Claim 5, dependent on Claim 1, is patentable by virtue of its dependency.

25

     Claim 6

         Claim 6, dependent on Claim 1, is patentable by virtue of its dependency.

     Claim 7

30   Claim 7, dependent on Claim 1, is patentable by virtue of its dependency.

Claim 8

Claim 8, dependent on Claim 1, is patentable by virtue of its dependency.

Claim 9

5   Claim 9 is patentable over the Syeda-Mahmood patent for the same reasons
presented above for Claim 1. That is, the Applicants submit that the Syeda-Mahmood
patent does not teach all of the elements of Claim 9, such as "using a first processing
element to convert electronic documents to files," "using said second processing element
to receive files and extract predetermined information," and "using a set of rules in said
10   reasoning element to extract metadata from the files."

The Applicants submit that the Examiner has not established a *prima facie* case of
anticipation for Claim 9 based on the teachings of the Syeda-Mahmood patent.
Therefore, Claim 9 is patentable over the Syeda-Mahmood patent.

15   Claim 10

Claim 10, dependent on Claim 9, is patentable by virtue of its dependency.

Claim 11

Claim 11, dependent on Claim 9, is patentable by virtue of its dependency.
20   Claim 11 is also patentable over the Syeda-Mahmood patent for the same reasons
presented above for Claim 3. That is, the Applicants submit that the Syeda-Mahmood
patent does not teach, disclose, or suggest "wherein said predetermined information is
substantially spatial layout facts" as is claimed in Claim 11.

25   Claim 12

Claim 12, dependent on Claim 9, is patentable by virtue of its dependency.

Claim 13

Claim 13, dependent on Claim 9, is patentable by virtue of its dependency.

30

Claim 14

Claim 14, dependent on Claim 9, is patentable by virtue of its dependency.

Claim 15

Claim 15, dependent on Claim 9, is patentable by virtue of its dependency.

5

Claim 16

Claim 16, dependent on Claim 9, is patentable by virtue of its dependency.

CONCLUSION

10   For the extensive reasons advanced above, the Appellants respectfully contend that each claim is patentable. Therefore, reversal of all rejections and objections is courteously solicited.

To the extent necessary, a petition for an extension of time under 37 CFR 1.136 is hereby made. Please charge any shortage of fees due in connection with the filing of this 15   paper, including extension of time fees, to deposit account no. 50-2691 and please credit any excess fees to such deposit account.

Respectfully submitted,

20

_7/11/05_
Date

Cary Tope-McKay
Registration No. 41,350

25

Cary Tope-McKay
TOPE-MCKAY & ASSOCIATES
23852 Pacific Coast Hwy. #311
Malibu, Ca 90265
30   Tel: 310.589.8158
Mobile: 310.383.7468
Fax: 310-943-2736
E-mail: cmckay@topemckay.com

Encl: Appendix A – pending Claims

35

CLAIMS

What is claimed is:

1. An apparatus for automatically extracting metadata from electronic documents comprising a first processing element, a second processing element, a reasoning element, and a database, wherein,

  i)   said first processing element is further configured to convert electronic documents into files;

  ii)   said first processing element is configured to provide the files to a second processing element;

  iii) said second processing element is configured to receive said files and extract predetermined information;

  iv)   said second processing element is further configured to provide said extracted predetermined information to said reasoning element;

  v)   said database is configured to also provide input to said reasoning element;

  vi)   said reasoning element is configured to use a set of rules to extract metadata from the files; and

  vii) said reasoning element provides an output of metadata.

2. An apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein said files are substantially format invariant data files such as Postscript files

3. An apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein said predetermined information is substantially spatial layout facts.

4. An apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein the second processing element and said database simultaneously input to the reasoning element.

5. An apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein said set of rules can be updated.

6. An apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein said metadata is substantially comprised of title, author, affiliation, author affiliation, and table of contents.

7. An apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein said metadata is provided to a user interface.

8. An apparatus for automatically extracting metadata from electronic documents as set forth in claim 1, wherein said metadata is provided to a storage medium.

9. A method for automatically extracting metadata from electronic documents providing a first processing element, a second processing element, a reasoning element, and a database and comprising the steps of:

    a) using said first processing element to convert electronic documents to files;

    b) further using said first processing element to provide the files to said second processing element;

    c) using said second processing element to receive said files and extract predetermined information;

    d) further using said second processing element to provide extracted predetermined information to said reasoning element;

    e) using said database to provide input to said reasoning element;

    f) using a set of rules in said reasoning element to extract metadata from the files;

    g) providing an out put of metadata from said reasoning element.

10. The method for automatically extracting metadata from electronic documents as set forth in claim **9**, wherein said files are substantially format invariant data files such as Postscript files.

11. A method for automatically extracting metadata from electronic documents as set forth in claim **9**, wherein said predetermined information is substantially spatial layout facts.

12. A method for automatically extracting metadata from electronic documents as set forth in claim **9**, wherein the second processing element and the database simultaneously input to the reasoning element.

13. A method for automatically extracting metadata from electronic documents as set forth in claim **9**, wherein said set of rules can be updated.

14. A method for automatically extracting metadata from electronic documents as set forth in claim **9**, wherein said metadata is substantially comprised of title, author, affiliation, author affiliation, and table of contents.

15. A method for automatically extracting metadata from electronic documents as set forth in claim **9**, wherein said metadata is provided to a user interface.

16. A method for automatically extracting metadata from electronic documents as set forth in claim **9**, wherein said metadata is provided to a storage medium.